

Участие в проектах LCG и EGEE

С.Д.Белов, В.В.Галактионов, В.Е.Жильцов, Н.С.Заикин, В.В.Кореньков,
Н.А.Кутовский, В.В.Мицын, Э.Г.Никонов, Д.А.Олейник, В.Д.Позе,

Т.А.Стриж, Е.А.Тихоненко, И.М.Ткачев

Лаборатория информационных технологий, ОИЯИ

Г.С.Шабратова

Лаборатория высоких энергий, ОИЯИ

Основной задачей проекта LCG (*LHC Computing Grid*; <http://lcg.web.cern.ch/LCG/>) является создание глобальной инфраструктуры региональных центров, расположенных в различных странах мира, для обработки, хранения и анализа данных физических экспериментов, строящихся на ускорителе Большой Адронный Коллайдер (*Large Hadron Collider, LHC*) в ЦЕРН. Новейшие технологии Грид являются основой построения этой инфраструктуры. Проект LCG осуществляется в две фазы: первая фаза (2001–2005 гг.) — создание прототипа и разработка проекта системы (LCG TDR [1]); вторая фаза (2005–2007 гг.) — создание инфраструктуры LCG, готовой к обработке, хранению и анализу данных на момент начала работы ускорителя в 2007 году. Следует отметить, что доступ к данным для всех экспериментов на ЛНС планируется осуществлять исключительно средствами LCG, т.е. каждый из участвующих в данных экспериментах институтов должен быть включен в инфраструктуру LCG, иметь при этом необходимый набор установленных LCG-сервисов, а сотрудники институтов (и, соответственно, ОИЯИ) должны быть зарегистрированными пользователями среды LCG и уметь работать в этой среде. Также надо подчеркнуть, что регистрация как оборудования и сервисов, так и пользователей среды LCG осуществляется в соответствии со строго определенными правилами.

Участие ОИЯИ в проекте LCG определяется трехсторонним Протоколом между ЦЕРН, правительством России и ОИЯИ, подписанным и одобренным в 2003 году. Согласно данному регламентирующему документу, сотрудничество по проекту LCG в ОИЯИ должно развиваться в следующих направлениях:

- создание и поддержка круглосуточной работоспособности инфраструктуры LCG в ОИЯИ;
- тестирование различных (внешних или разработанных внутри LCG) программных средств Грид, принятых для внедрения в инфраструктуру LCG;
- оценка технологий Грид (в частности, набора инструментальных средств Globus toolkit (<http://www.globus.org/>) и программного обеспечения (ПО) Грид из коммерческой сферы (<http://gridcafe.web.cern.ch/gridcafe/gridprojects/industry.html>) в контексте их возможного использования в проекте LCG;
- разработка и поддержка репозитория физических генераторов и базы данных модельных физических событий.

Помимо вышечисленных задач, на 2004–2005 годы также было запланировано развитие и поддержка системы CASTOR — иерархической системы управления хранением данных, разработанной в ЦЕРН (*CERN Advanced STORAGE manager*; <http://castor.web.cern.ch/castor/>), и участие в проекте ARDA по созданию системы распределенного анализа для ЛНС (*A Realization of Distributed Analysis for LHC*; <http://lcg.web.cern.ch/LCG/activities/arda/arda.html>).

Другим крупнейшим современным проектом по созданию глобальной грид-инфраструктуры является проект EGEE (*Enabling Grids for E-science*; <http://www.eu-egee.org>). Проект EGEE стартовал 1 апреля 2004 года. Он рассчитан на два года и задуман как часть четырехлетней программы. ОИЯИ и ряд российских научных центров, успешно зарекомендовавших себя в проектах EU DataGrid (<http://www.eu-data-grid.org>) и LCG, были приглашены европейскими партнерами участвовать в проекте EGEE. Следует подчеркнуть, что главным результатом участия ОИЯИ и российских институтов в проекте EU DataGrid явилось получение опыта работы с новейшим программным обеспечением Грид и включение российского сегмента в европейскую инфраструктуру EU DataGRID.

Цель проекта EGEE — объединить уже ведущиеся национальные, региональные и тематические грид-разработки в единую цельную грид-инфраструктуру для поддержки научных исследований. EGEE предоставит круглосуточный доступ к самым высокопроизводительным вычислительным ресурсам независимо от их географического положения. Пользоваться инфраструктурой смогут географически распределенные научные сообщества, имеющие потребность в общих для них вычислительных возможностях и готовые объединить свои собственные вычислительные инфраструктуры в соответствии с принципами общего доступа.

На 4-ой конференции EGEE в Пизе в октябре 2005 года были доложены основные результаты, достигнутые на пути построения грид инфраструктуры для научных целей, а также отмечено, что в рамках инфраструктуры EGEE, которая состоит более чем из 150 сайтов в Европе, Америке и Азии, уже было обработано более двух миллионов заданий, что эквивалентно примерно тысяче годам процессорного времени одного компьютера.

Инфраструктура EGEE строится на базе исследовательской сети EC GEANT. В ней используется опыт, накопленный в европейских проектах DataGrid, Datatag и Crossgrid, в национальных проектах: например, e-Science (Великобритания), INFN Grid (Италия), Nordugrid (скандинавские страны) и в американском проекте Trillium. Инфраструктура EGEE обеспечит возможность совместной работы с другими Грид-системами во всем мире, включая США и Азию, что будет способствовать установлению всемирной инфраструктуры Грид.

Функциональными особенностями проекта EGEE являются:

- упрощенный доступ как уменьшение издержек, обусловленных разнообразием не связанных между собой систем учета пользователей; пользователи смогут объединяться в виртуальные организации с доступом к Grid — инфраструктуре, содержащей нужные каждому пользователю рабочие ресурсы в рамках соответствующей виртуальной организации;
- выполнение вычислений по требованию с эффективным распределением ресурсов и, соответственно, сокращением времени доступа к ним;
- надежный и быстрый, географически широко распределенный доступ к ресурсам при хорошей совместимости сетей;
- исключительно большой объем ресурсов, что может быть достигнуто только благодаря согласованности этих ресурсов и прикладных пользовательских сообществ; такой объем ресурсов не может предоставить ни один компьютерный центр;

- совместное использование ПО и данных благодаря единой структуры вычислительных средств;
- необходимый уровень поддержки приложений, обеспечиваемый высокой компетентностью всех участников проекта EGEE.

В настоящее время на всех сайтах, участвующих в проекте EGEE, установлено ПО LCG-2, и, таким образом, можно определить инфраструктуру проекта как единую инфраструктуру LCG/EGEE.

ОИЯИ и российские институты, участвующие в EGEE, образовали консорциум РДИГ *Российский Грид для интенсивных операций с данными* — *Russian Data Intensive Grid, RDIG*) для эффективного выполнения работ по проекту и развитию в России инфраструктуры EGEE, с вовлечением на следующих этапах проекта других организаций из различных областей науки, образования и промышленности. Факт создания консорциума РДИГ зафиксирован в Меморандуме о создании в России компьютерной инфраструктуры Грид, подписанным в 2003 году директорами следующих организаций: ОИЯИ (Объединенный институт ядерных исследований, Дубна), ИФВЭ (Институт физики высоких энергий, Протвино), ИМПБ РАН (Институт математических проблем биологии РАН, Пущино), ИТЭФ (Институт теоретической и экспериментальной физики, Москва), ИПМ РАН (Институт прикладной математики РАН, Москва), ПИЯФ (Петербургский институт ядерной физики, Гатчина), РНЦ КИ (Российский научный центр "Курчатовский институт", Москва), НИИЯФ МГУ (Научно-исследовательский институт ядерной физики МГУ, Москва).

В настоящее время консорциум RDIG является частью инфраструктуры LCG/EGEE. ОИЯИ участвует в следующих направлениях деятельности EGEE: SA1 — системная поддержка функционирования и управление Грид-инфраструктурой; NA2 — информационная служба и распространение знаний; NA3 — обучение; NA4 — идентификация и поддержка программных приложений.

К середине 2004 года в ОИЯИ была создана инфраструктура LCG/EGEE как составная часть Центрального информационно-вычислительного комплекса (ЦИВК) ОИЯИ и паневропейской инфраструктуры Грид. В ОИЯИ эта инфраструктура состоит из комплекса управляющих серверов и 20 вычислительных узлов. Для доступа пользователей к глобальной инфраструктуре Грид на одной из интерактивных машин ЦИВК ОИЯИ (*krub03.jinr.ru*) реализован сервис User Interface. На LCG инфраструктуре ОИЯИ также установлены следующие сервисы:

- доступ к ресурсам хранения информации (Storage Element; SE) на основе системы DPM;
- Грид-очередь пакетной обработки заданий (Computing Element; CE);
- сервис поиска ресурсов (Resource Broker; RB);
- информационный сервис (BDII);
- расширенный сервис для предоставления разрешения на работу в среде Грид (MyProxy).

Первые два из перечисленных выше сервисов являются обязательными для обеспечения функционирования инфраструктуры LCG. Последние три сервиса поставлены дополнительно. Такое дублирование сервисов повышает надежность функционирования RDIG инфраструктуры в целом. На протяжении 2004–2005 гг. дважды

обновлялись версии ПО LCG: осуществлялся переход на LCG 2.4 и LCG 2.6. Также был произведен переход на операционную систему Scientific Linux.

Для мониторинга российских LCG-сайтов в ОИЯИ установлен и поддерживается отдельный сервер <http://rocmon.jinr.ru:8080> [2]. В настоящее время мониторинг базируется на системе MonALISA. Параллельно ведутся освоение и изучение возможностей использования других систем (GridICE, MapCenter).

К настоящему времени в рамках проекта LCG совместно с российскими и зарубежными коллегами выполнено:

- развитие системы G2G(GoToGrid) для web-ориентированной поддержки инсталляции и настройки сайтов LCG (разработан комплекс модулей для создания конфигурационных файлов, которые используются непосредственно при инсталляции узлов, а также при генерации руководств для администраторов по установке и настройке) [3];
- развитие системы CASTOR2 (разработка служебного модуля для выполнения функции сборщика мусора в системе);
- разработка структуры базы данных моделированных Монте-Карло событий в CERN; создание набора базовых модулей; разработка WEB-интерфейса для создания/добавления в базу данных новых данных (описания файлов с событиями и связанных с ними объектов) (<http://mcdm.cern.ch>) [4];
- проведение тестов по проверке надежности передачи данных при помощи протокола GridFTP, реализованного в пакете Globus Toolkit 3.0 [5].

Ключевым моментом в грид-компьютинге является ПО промежуточного слоя, которое организует и интегрирует различные вычислительные ресурсы и обеспечивает пользователям единый интерфейс к гриду. В рамках проекта EGEE по направлению SA1 ведутся работы по системной поддержке собственно инфраструктуры LCG/EGEE, осуществляется мониторинг ресурсов консорциума RDIG, проводится тестирование ПО промежуточного слоя и координация проведения сеансов Service Challenge по проверке работоспособности инфраструктуры RDIG при интенсивной передаче данных в европейские центры [6]. В рамках этого направления выполнено:

- тестирование каталога метаданных и комбинированного каталога файлов и реплик ПО промежуточного слоя EGEE — gLite;
- разработка программы постоянного мониторинга WMS (Workload Management System) сайта INFN gundam.cnaf.infn.it на испытательном полигоне (testbed) для gLite;
- оценка сервисов и аспектов безопасности, авторизации, размещения ресурсов, а также взаимодействия с системой управления загрузки (WMS) gLite для новой программной платформы грид - системы ОМИ (Open Middleware Infrastructure Institute)(в сотрудничестве с ИПМ) [7];
- оценка набора инструментальных средств Globus Toolkit 4 (GT4) (в сотрудничестве с ИПМ и НИИЯФ МГУ): инсталляция и конфигурирование, оценка сервисов и аспектов безопасности и авторизации, интерфейсы JAVA WS-Core, WS-GRAM, GridFTP, RLS, RFT , WS-MDS4, сравнение соответствующих компонент GT4 и gLite [8].

В рамках NA2 осуществляется координация работ по этому направлению во всем консорциуме RDIG, создание информационного портала RDIG (<http://www.egee-rdig.ru>), организация конференций, распространение знаний о Grid в странах-участницах ОИЯИ [9].

Информационный портал RDIG представляет собой комплекс web-сайтов и баз данных для создания и поддержки единого информационного пространства в рамках проекта EGEE RDIG. Портал реализован на web-сервере Apache 2 с использованием СУБД Oracle и php — Personal Home Page [10].

В рамках NA3 проводились обучающие курсы и стажировки различного уровня: как для начинающих пользователей Грид, так и для системных администраторов. В течение 2004–2005 годов были проведены вводные курсы по работе в Грид-среде для сотрудников ОИЯИ и участников конференции “Распределенные вычисления и Грид-технологии в науке и образовании”, стажировка для системных администраторов российских институтов и стран-участниц ОИЯИ, а также специализированные курсы для участников эксперимента CMS.

В рамках NA4 осуществлялась координация данного направления в RDIG, установка и тестирование специализированного программного обеспечения для экспериментов ALICE, ATLAS, CMS и LHCb в среде LCG-2, участие в сеансах Data Challenge, привлечение пользователей в виртуальные организации ALICE, ATLAS, CMS и LHCb, создание моделей компьютинга для экспериментов LHC в ОИЯИ и России [11,12].

Создание в ОИЯИ инфраструктуры LCG, включение этой инфраструктуры в глобальные европейскую и мировую инфраструктуры и соответствующее участие ОИЯИ в проектах LCG и EGEE дает возможность всем сотрудникам ОИЯИ уже сегодня и сейчас стать пользователями глобальной компьютерной инфраструктуры, а для участников экспериментов на LHC создается фундамент их дальнейшего полноценного участия в экспериментах ALICE, ATLAS и CMS на действующей стадии данных экспериментов.

Список литературы

- [1] LHC Computing Grid Technical Design Report, LCG-TDR-00, CERN-LHC-2005-024, CERN, Geneva, 20 June 2005, available at <http://cern.ch/lcg/tdr>.
- [2] С.Д. Белов, И.М. Ткачев, Система мониторинга российской Грид-инфраструктуры, в данном сборнике, с.46.
- [3] Z.Diez-Andino, O.Keeble, P.Nyczyk, A.Retico, M.Shultz, I.Tkachev, A.Usai, GoToGrid - A Web-Oriented Tool in Support to Sites for LCG Installations, in the Proceedings of the International Conference on Computing in High-Energy Physics: CHEP'04, Interlaken, Switzerland, 2004, CERN, 2004, vol.2,pp.1018–1021.
- [4] P.Bartalini et al., LCG Generator, in the Proceedings of the International Conference on Computing in High-Energy Physics: CHEP'04, Interlaken, Switzerland, 2004, CERN, 2004, vol.2,pp.183–186; L. Dudko et al., hep-ph/0404241.
- [5] A.Demichev, D.Foster, V.Kalyaev, A.Kryukov, M.Lamanna, V.Pose, R.Rocha, Experiences with the gLite Grid Middleware, in the Proceedings of International

- conference YDistributed Computing and GRID-technologies in Science and Education, Dubna, 2004, CERN, 2004, vol.2,pp.834–837.
- [6] V.Jiltsov, V.Korenkov, JINR and Russia in LCG Service Challenge Activity, in the Book of Abstracts of the XXth international Symposium on Nuclear Electronics and Computing, Dubna, 2005.
- [7] P.Berezovsky, E. Huhlaev, V. Kovalenko, V.Galaktionov, N.Kutovskiy, V.Pose, D.Semyachkin, I.Tkachev, OMI distribution evaluation activity, в данном сборнике, с.55.
- [8] Галактионов В.В., Кутовский Н.А. Тестирование GT4 в ОИЯИ, в данном сборнике, с.48.
- [9] T.Strizh, Grid Awareness Dissemination and Outreach Activities in Russian Federation, in the Book of Abstracts of the XXth international Symposium on Nuclear Electronics and Computing, Dubna, 2005; T.Strizh “NA2 in Russia”, EGEE Second Conference, 2004, Hague, <http://agenda.cern.ch/askArchive.php?base=agenda&categ=a045128&id=a045128s1t8%2Ftransparencies%2Fna2Russia.pdf>; T.Strizh “Na2 in Russia - future plans” 2005, Athens, Greece, The Third EGEE Conference <http://indico.cern.ch/materialDisplay.py?contribId=12&sessionId=70&materialId=slides&confId=0513>.
- [10] Ilyin V.A., Korenkov V.V., Nikonov E.G., Olynik D.A., Maksimiva Yu.N., Prikhodko A.V., Semenov R.N. , Isayev R.A. Web-portal www.egee-rdig.ru: unified information environment of RDIG participants. In the Proceedings of International conference "Distributed Computing and GRID-technologies in Science and Education", Dubna, 2004,JINR,D11-2004-205 pp. 116 – 123.
- [11] E.Tikhonenko, Application Identification and Support (NA4) Activities in the RDIG-EGEE consortium , in the Book of Abstracts of the XXth international Symposium on Nuclear Electronics and Computing, Dubna, 2005.
- [12] V.Gavrilov, I.Golutvin, V.Ilyin, O.Kodolova, V.Korenkov, S.Shmatov, E.Tikhonenko, RDMS CMS Computing Model, in the Proceedings of International conference "Distributed Computing and GRID-technologies in Science and Education", Dubna, 2004In: Proc. of Int. conference “Distributed Computing and Grid-Technologies in Science and Education”, June 29 – July 2, 2004, Dubna, Russia,JINR,D11-2004-205 pp. 240 – 247.